DOCUMENT RESUME

ED 274 149 FL 015 909

AUTHOR Wimbish, John
TITLE The Languages of the Zambales Mountains: A Philippine Lexicostatistic Study.
PUB DATE 86
NOTE 11p.; In: Work Papers of the Summer Institute of Linguistics, Volume 30, 1986; see FL 015 901.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Computer Oriented Programs; *Computer Software; Foreign Countries; Language Research; *Language Typology; Lexicology; *Statistical Analysis; *Structural Analysis (Linguistics); Surveys; Uncommonly Taught Languages; *Vocabulary
IDENTIFIERS Philippines; Zambales Mountains

ABSTRACT

A survey gathering word lists from 40 locations in the Zambales Mountains of the Philippines, for statistical comparison, resulted in the definition of six distinct Negrito languages: Abelen, Magqanchi, Aberlen, Magqindi, Ambala, and Mabeken. With the use of a computer program, WORDSURV, designed for word list comparison, an Ayta language tree was generated based on shared cognate percentages. The language tree identified the locations of each of the six languages and four trade languages (Sambal, Kapangan, Tagalog, and Ilocano). (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. Derbyshire

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED274149

THE LANGUAGES OF THE ZAMBALES MOUNTAINS: A Philippine Lexicostatistic Study

John Wimbish

1 Overview
2 Introduction
3 Data gathering and analysis
4 The Zambales Ayta Language Tree
5 Summary
Appendix
Notes
References

1 Overview

A lexicostatistic survey was conducted in early 1985 as part of a comprehensive study of the languages spoken in the Zambales Mountains of the Philippines. The survey involved the collection and analysis of some forty word lists, resulting in the definition of six distinct languages in an area whose linguistic situation had previously been relatively unknown. This paper presents these survey findings in the form of a language tree, and introduces a general-use computer program which greatly enhanced the accuracy and speed of the lexicostatistic analysis.

2 Introduction

A yet unresolved mystery in anthropological studies has been the history of the Philippine Negrito community, a people of dark complexion, short stature, and black, curly hair. Except for their preference of dwelling in the more mountainous areas of the country, the only trait separating these people from other Filipinos is their physical makeup. No cultural bond exists between them, nor are there any linguistic commonalities. Rather, their languages and customs tend to be related more to those of their geographic neighbors. All attempts to find such characteristics to unite the different Negrito groups and to distinguish them from the lowland communities have ended in frustration (Headland 1975:49-50).

The Zambales mountains (Figure 1) are home for some ten thousand Negritos. These Zambales Negritos

FL 015 909

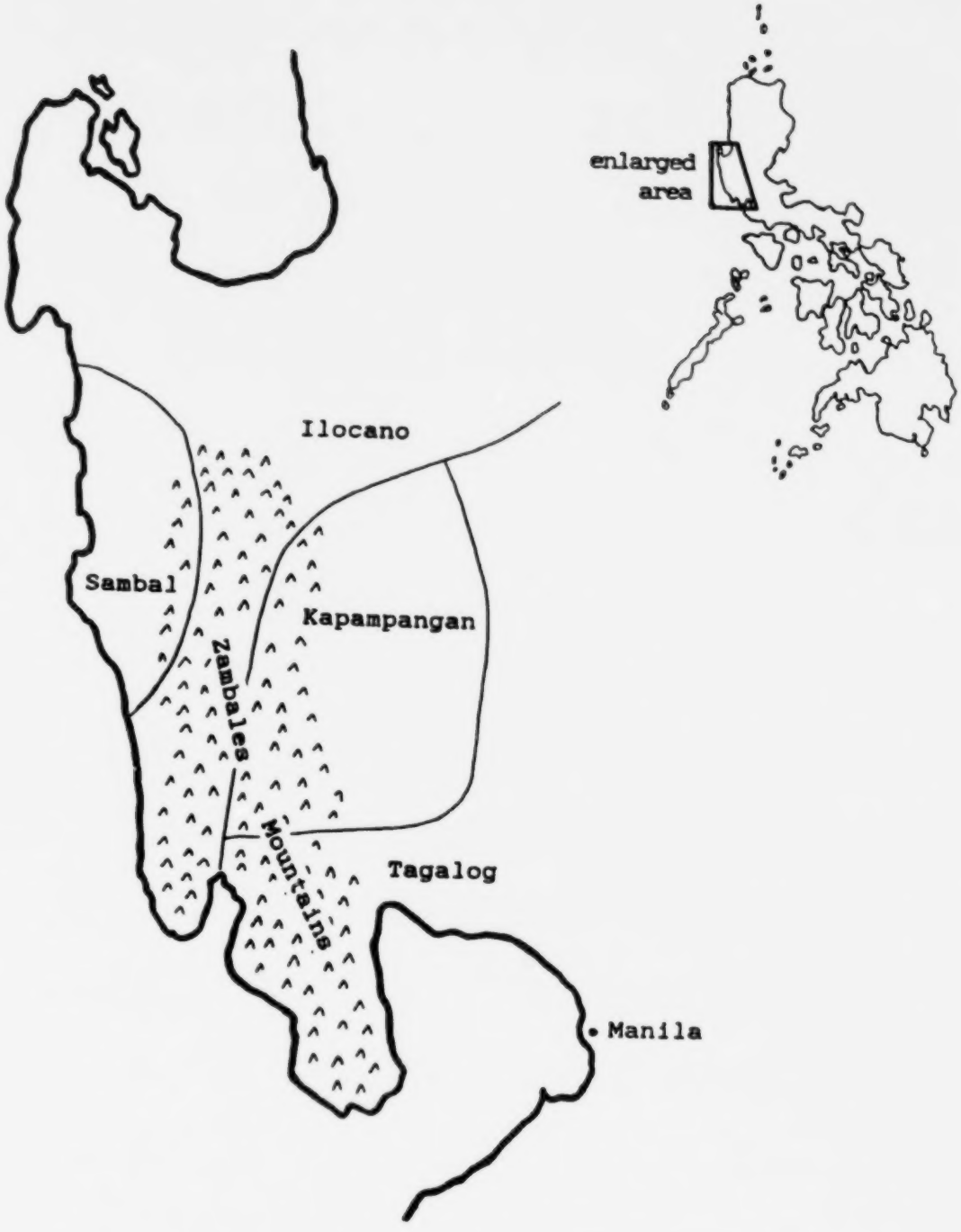


Figure 1 - Location of Zambales Mountains
and Trade Languages

(alternately, Aytas) were once suspected to all be speakers of a language closely related to Sambal, a nearby lowland language (Busenitz 1973). Later studies (Kyle 1977, Savage 1982) revealed the existence of several Ayta dialects, but due to limits on the number of locations they were able to visit, the geographic boundaries of these dialects remained undefined. The need for a comprehensive survey was established when attempts to use materials developed in the Sambal language for literacy instruction met with comprehension difficulties among the Aytas; the teachers being informed that there were nearly as many Ayta dialects as villages, and that few Ayta could understand Sambal.

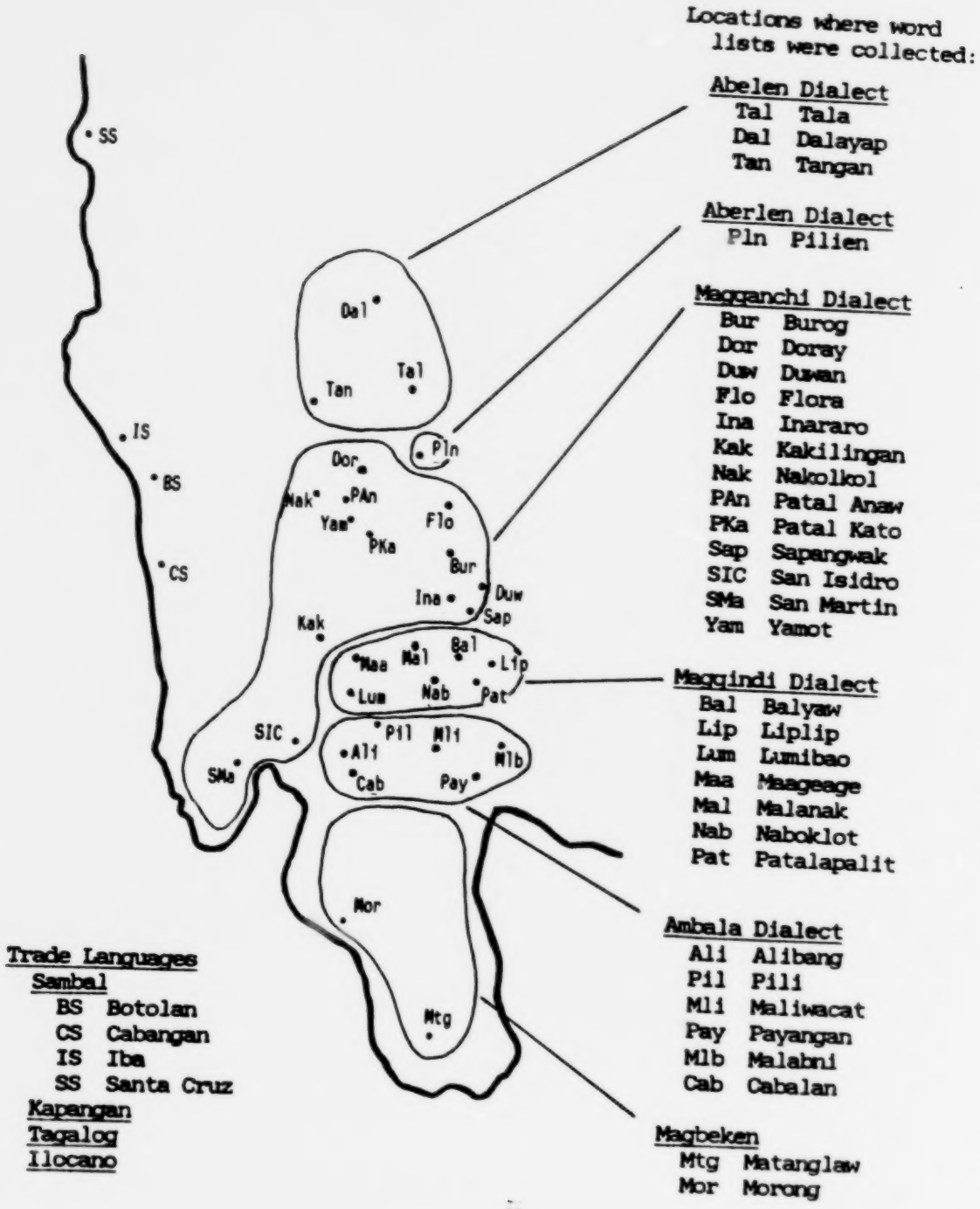
With the goal of developing adequate language materials for all of the people of the Zambales mountains, the Summer Institute of Linguistics (SIL) planned a two-part survey. The first phase would be a lexicostatistical study, entailing the gathering and comparison of word lists from every location possible, in order to gain a picture of the overall language situation. The second phase would then consist of Casad-type text testing and sociolinguistic study, to further define each language's geographic boundaries and to indicate the Ayta's understanding and use of the trade languages in each area. What follows is a presentation of the results of the initial phase.

3 Data gathering and analysis

The collection of word lists has often been the first step in linguistic survey procedure, providing a general and rough classification of the dialects in the target area. Experience among SIL members in Philippine comparisons has generated two empirical definitions for the interpretation of cognate scores between lists[1]. The first states that two lists represent the exact same dialect if their vocabularies are greater than 95 percent cognate. The second claims that two lists represent mutually unintelligible languages if their vocabularies are less than 60 percent cognate[2]. Lists with percentages between the two figures can only be tentatively classified, relying on text testing to give a more accurate assessment of the intelligibility between the two locations.

The Reid Word List, consisting of 372 items tailored for use in the Philippines, was collected in 40 locations[3], as represented in Figure 2. The comparison of these lists created a challenge: the mathematics predicted nearly eight

Figure 2 - Ayta Dialect Boundaries



hundred comparisons, requiring over six months of man-time to perform. As an alternative, the computer program WORDSURV[4] was developed.

WORDSURV allows the comparison of up to ninety word lists. Each list may contain as many as 400 items. A list is entered once through the program's editor, during which time the analyst groups the elicited forms into cognate sets with forms from other lists he has previously entered. Once a list has been entered, its cognate percentages with the other lists on file may be computed. This computation requires only a few minutes of computer time.

The analyst then arranges the cognate percentages into matrix form, not unlike distance charts on road maps. In the sample matrix in Table 1, the dialect spoken in the village of Doray shows 93 percent of its basic vocabulary to be cognate with the dialect spoken in Burog. Dialect names are written in two sections, showing language grouping and location: thus Magqanchi-Doray refers to the variety of the Magqanchi language spoken at Doray.

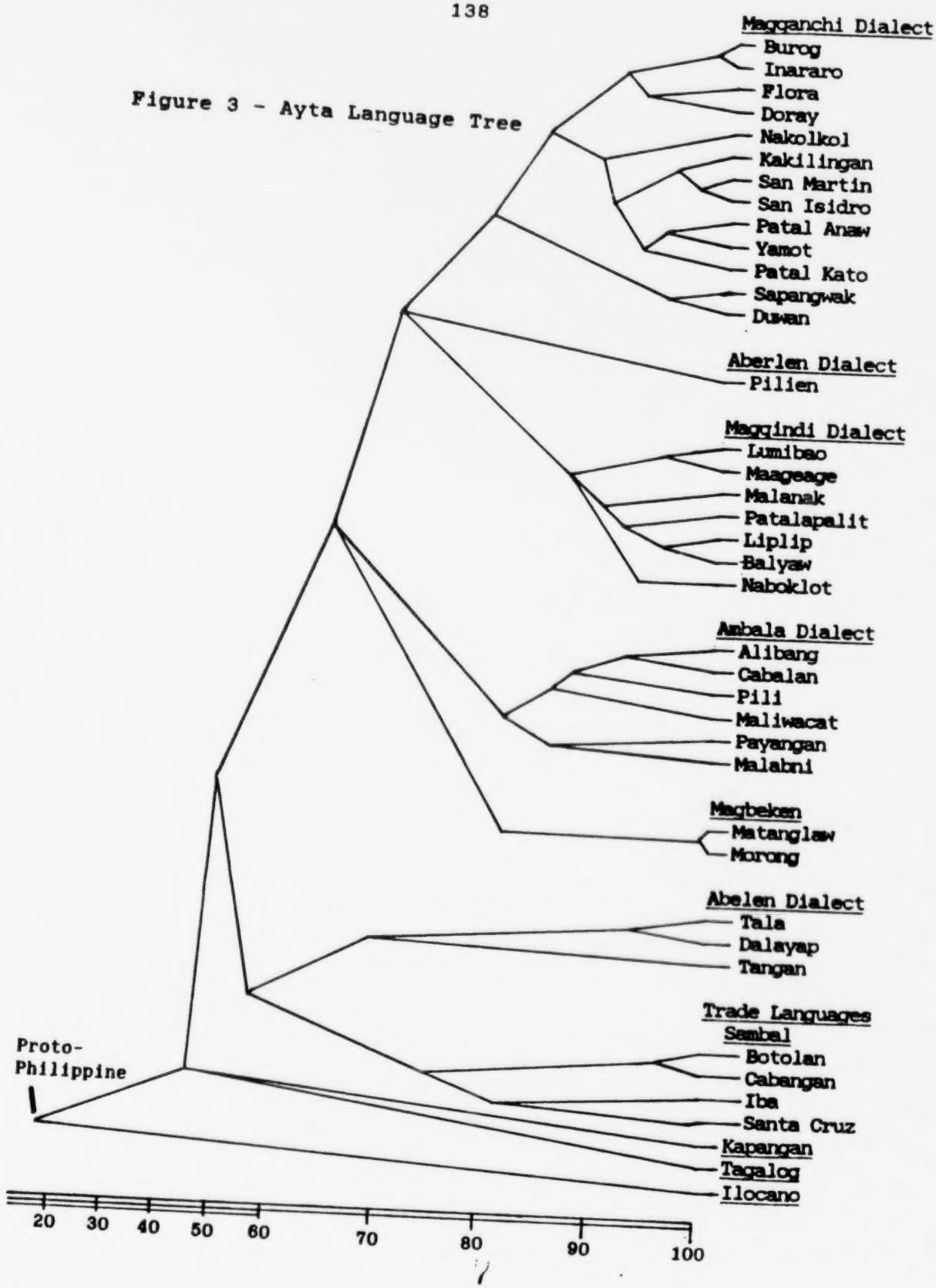
Magqanchi-Doray
92 Magqanchi-Flora
93 93 Magqanchi-Burog
90 90 98 Magqanchi-Inaararo
89 86 88 88 Magqanchi-Nakolkol

Table 1 - A Sample Matrix

Several methods of deriving a language tree from a cognate matrix are discussed by Walton (1977), who found that the "lowest-number" method generated the most accurate results in his study of 122 Philippine languages. This method, combined with the author's knowledge of geographic and lexical borrowing patterns in the area, was employed in the analysis.

It is recognized that a derivation based on shared cognate percentages lacks much of the accuracy of the tree that would be produced through a rigorous application of the comparative method. But it is also recognized that such an application of the comparative method with these word lists would require an extensive time investment. The tree presented in this paper is intended to merely provide a rough picture of the development of these Ayta languages, with the understanding that further study would undoubtedly generate refinements.

Figure 3 - Ayta Language Tree



4 The Zambales Ayta Language Tree

The circled areas of Figure 2 indicate the proposed major groupings derived from the word list comparisons. The six languages defined are:

Ayta Language	Population[5]
1. Mag-anchi	4166
2. Aberlen	184
3. Mag-indi	2483
4. Ambala	1657
5. Mabeken	572
6. Abelen	2000

TABLE 2 - Major Ayta Language Groupings

It is anticipated that dialects from villages within a given language grouping will exhibit a fair amount of intelligibility; village dialects separated by these language boundaries should prove inherently unintelligible.

The tree derived from the cognate matrix is presented in Figure 3. The cognate percentages are listed in the appendix.

5 Summary

The Zambales Mountains contain - different Negrito languages, Abelen, Magqanchi, Abe , Magqindi, Ambala, and Mabeken. An assessment of the ility of the speakers of these languages to understand their neighbors and the trade languages is the goal of the second phase of the survey. Though one cannot predict intelligibility on the basis of cognate scores alone, the groupings presented should provide a starting point for further examination of language use patterns among these Zambales Ayta.

APPENDIX

AYTA LEXICOSTATISTIC MATRIX

Accuracy codes for each word list are given in brackets; see note 3 for an explanation of their meaning. The cognate percentage between two lists is read as the intersection between their names. For example, the Abelen-Dalayap and Magqanchi-Doray lists are 63 percent cognate.

(C) Sambal-Iba
(C) 81 Sambal-Santa Cruz
(B) 75 74 Sambal-Cabanggan
(A) 76 74 96 Sambal-Botolan
(B) 44 44 45 46 Ilocano
(B) 41 44 44 44 47 Pangasinan
(B) 48 47 54 53 45 38 Tagalog
(C) 44 43 48 47 35 34 53 Pasonge
(C) 57 56 64 66 38 38 41 37 Abelen-Dalayap
(B) 59 57 66 68 38 40 44 38 93 Abelen-Iala
(C) 60 59 68 73 38 40 45 42 88 78 Abelen-Tangan
(B) 60 59 73 76 36 37 50 48 59 61 65 Magqanchi-Flora
(A) 60 59 75 78 40 37 50 49 63 62 67 83 Magqanchi-Burog
(A) 60 59 75 78 40 38 49 47 63 63 67 90 98 Magqanchi-Inaoro
(B) 66 64 79 82 42 39 51 51 63 63 69 92 93 80 Magqanchi-Doray
(B) 65 64 82 86 40 39 51 47 65 66 72 86 88 88 88 Magqanchi-Nakolkol
(B) 67 80 84 88 41 42 51 46 66 68 72 85 88 89 87 88 Magqanchi-Kaklitingan
(A) 64 64 83 85 40 40 50 47 65 67 72 83 86 87 87 88 95 Magqanchi-San Martin
(A) 65 64 83 87 41 41 51 47 65 67 72 85 87 87 88 90 95 97 Magqanchi-San Isidro
(A) 58 58 88 73 36 36 48 50 60 61 64 88 91 88 87 84 83 81 83 Magqanchi-Sapangpak
(B) 55 63 55 69 34 34 47 51 53 56 60 83 85 82 83 80 78 78 79 84 Magqanchi-Quezon
(B) 66 64 80 85 40 42 53 48 58 64 71 86 88 87 88 90 90 92 84 84 79 Magqanchi-Petal Anew
(B) 66 65 80 85 38 41 52 47 80 65 72 86 90 88 90 91 88 88 84 85 81 84 Magqanchi-Yaapt
(B) 63 61 76 82 37 39 49 47 58 63 70 87 88 87 87 88 88 89 91 84 78 82 93 Magqanchi-Patalleeto
(B) 65 64 80 83 42 42 50 48 64 66 75 78 77 76 84 82 81 81 82 75 70 81 83 80 Abarlen-Pitlen
(B) 53 53 65 69 33 34 42 44 56 57 61 77 78 77 78 78 78 76 77 83 76 77 79 77 70 Magqindi-Patalapalit
(B) 57 57 68 73 36 37 47 47 61 62 66 81 83 82 83 83 83 82 83 90 85 81 82 81 72 81 Magqindi-Balyaw
(B) 55 53 66 70 35 36 45 45 59 58 64 78 81 78 80 78 Bo 78 79 85 79 79 80 79 70 90 95 Magqindi-Lipitp
(B) 55 54 63 66 34 35 44 46 58 58 62 77 80 78 78 77 78 75 76 84 78 77 79 78 70 89 89 87 Magqindi-Malanak
(B) 55 54 67 70 37 38 45 45 80 59 65 75 78 78 77 78 83 77 79 82 78 78 80 77 71 88 88 89 85 Magqindi-Luaibao
(B) 58 57 70 73 37 38 45 45 61 62 66 77 81 80 79 80 85 81 80 84 77 80 82 80 73 88 88 87 86 96 Magqindi-Masgeaga
(B) 54 53 64 67 33 35 46 49 55 56 61 76 76 75 75 75 74 73 74 84 77 74 76 75 69 88 87 83 84 82 82 Magqindi-Nabaklat
(B) 61 58 72 74 39 38 55 48 56 56 62 72 73 72 74 74 77 74 75 73 69 75 75 74 70 88 74 70 89 70 71 71 Aeta-Alitang
(B) 63 61 77 80 42 42 54 50 80 62 67 76 78 78 79 80 82 81 81 79 74 82 82 80 76 78 82 77 75 80 81 80 90 Aeta-Pili
(C) 58 57 69 70 38 38 54 47 56 55 57 67 67 66 68 68 70 67 67 67 64 67 67 65 66 64 88 85 84 67 88 67 86 85 Aeta-Maliasat
(B) 60 57 68 70 39 38 64 52 54 55 58 67 68 67 88 72 70 67 88 68 64 88 88 67 88 85 89 67 86 67 68 69 86 83 83 Aeta-Payangan
(B) 61 57 70 71 38 37 54 52 54 55 58 71 88 88 72 70 71 70 71 70 89 70 70 70 88 85 89 67 86 86 67 70 86 83 81 85 Aeta-Malabni
(B) 58 57 69 71 37 37 55 50 55 55 59 69 70 69 70 71 72 70 71 72 69 71 72 70 86 88 73 70 69 70 70 71 82 87 87 87 87 Aeta-Cabalan
(B) 56 55 63 63 40 37 63 49 51 51 55 61 61 61 61 64 63 61 63 80 57 64 64 61 61 54 56 55 56 55 56 56 87 85 68 73 65 70 Magbalen-Morong
(B) 53 51 59 80 38 35 61 47 49 49 53 58 58 58 59 80 80 69 80 58 55 62 61 58 58 55 58 57 58 57 59 58 70 89 72 75 68 73 99 Magbalen-Metangleo

Notes

1. The word "cognate" in comparative linguistics normally refers to language forms which have descended from a single proto-form. The surveyor does not perform the rigorous analysis necessary to determine true cognates. Instead, he groups elicited forms into sets based on their similarities and his knowledge of the language family. Cognate, as used in survey and in this paper, should be understood as referring to these quickly determined "apparent" cognates.

2. These empirical figures are rules of thumb used by SIL surveyors in the Philippines. Simons (1983:67-100) explores several models which use lexicostatistics to predict intelligibility, and from which such figures might be derived.

3. Lawrence Reid developed the word list in 1966 by combining the relevant terms from the Swadesh 200 list with concepts central to Philippine life. The list has been used extensively by members of SIL in conducting language surveys, and is reproduced by Reid (1971) with data from some forty Philippine languages.

Most of the word lists were collected by Kurt and Margarette Storck, though a few were elicited or checked by the author. Word lists which had been obtained from previous surveys were rechecked. The accuracy codes in the appendix reflect the checking. A list was given the grade [C] when first obtained (either from archives or elicitation). A thorough recheck merited an upgrade to [B]. A grade of [A] indicates that the list was subjected to a second recheck. The compilation of multiple forms on the computer provided opportunity to recognize, and therefore recheck, suspicious items.

4. The WORDSURV program is available in the Summer Institute of Linguistics software library in the International Linguistic Center in Dallas, Texas. Several similar programs have been developed for use with small lap-top computers (Simons 1984), but these lack the full screen editing capabilities of WORDSURV, and can only handle a few word lists at a time. WORDSURV is only available for IBM-compatible computers and requires a 25-line display.

5. Population figures for the first five languages listed are from Storck 1986, as taken from government census. The figure for Abelen is an estimate taken from Wimbish 1984.

References

Busenitz, Bob. 1973. Sambal Survey. Ms. Summer Institute of Linguistics. Manila.

Casad, Eugene H. 1974. Dialect intelligibility testing. Norman, Oklahoma. Summer Institute of Linguistics.

Headland, Thomas N. 1975. Report of eastern Luzon language survey. Philippine Journal of Linguistics. Vol. 6, No. 2. Manila.

Kyle, Marc. 1977. Sambal Survey Report. Ms. Summer Institute of Linguistics. Manila.

Reid, Lawrence A. 1971. Philippine minor languages: word lists and phonologies. Oceanic linguistics special publication no. 8. University of Hawaii Press.

Savage, Dale. 1982. Pre-report memo -- Sambal / Ayta survey. Summer Institute of Linguistics. Manila.

Simons, Gary F. 1983. Language Variation and Limits to Communication. Summer Institute of Linguistics. Dallas.

-----. 1984. Word list analysis with a notebook computer. Occasional Publications in Academic Computing. No. 5. Summer Institute of Linguistics. Dallas.

Storck, Kurt and Margarette. 1986. A sociolinguistic survey report of the Ayta of the Western Cordillera of Luzon Island, Philippines. Ms. Summer Institute of Linguistics.

Walton, Charles. 1977. A Philippine language tree. Presented at the Austronesian Symposium. University of Hawaii. Honolulu.

Wimbish, John S. 1984. Abelen survey report. Ms. Summer Institute of Linguistics.